CMSC416: Introduction to Parallel Computing

Topic: Parallel Networks and File Systems
Date: April 18th, 2024

Parallel Networks:

- **Overview**: Parallel networks in High-Performance Computing (HPC) clusters facilitate simultaneous data processing and communication across multiple nodes. This architecture significantly boosts the system's efficiency and performance.
- **What we will do:** Figure out what parallel networks look like especially on HPC clusters.

Filesystems:

- We will discuss file systems like the directory we use on Zaratan.
- On Zaratan there is a separate set of nodes on which the file system is virtually mounted.
- This will be covered more in the following lecture

High-speed interconnection networks:

- **Locations**: HPC systems are on premise systems so there's a physical area where the server/datacenter exists in a single area. These systems are much faster than cloud systems that are spread out across multiple regions.
- **Benefits**: Offers lower latency and higher data transfer speeds due to the proximity of hardware resources.
- **Network Topology**: The structural arrangement of network elements. The most traditional topology in HPC settings is the Fat-tree topology.
- The oldest network topology in use is called a Fat-tree.

Network Components:

- **Network interface card (NIC):** Acts as a bridge between the CPU and the network, transmitting data from the CPU's memory to the network. Connects the processor or cpu to the network. On the cpu chip, this takes data from the memory of the cpu and copies the data onto the cable of the network. So it interfaces the cpu and the network.
- **Router or Switch:** Devices that connect multiple CPUs or routers, forming a network. These are critical for scaling the network infrastructure.This helps us create a network of lots of cpu’s or routers. This helps us build a large network of lots of components
- **Network Cables:** Comprise either copper (short distance, less cost) or fiber optic (long distance, higher cost) materials.Copper or fiber optic cables are the 2 kinds. Optical cables are newer and different in cost.

- **Network Topology:** The specific arrangement used to connect routers, which influences performance and efficiency. Another way of looking at it is the way you connect routers together is called the network Topology

Key Definitions:

- **Hops/Distance:** Everytime you jump from one router to another router that is 1 hop, going from a node to a router is also 1 hop. So how far you have to travel between nodes is the distance between the nodes. You can also include the jumps between the NIC and the router.
- **Network Diameter**: Is defined as the length of the shortest path between the farthest nodes of the network. Aka it is the longest shortest path between any two nodes in the network. This is crucial for understanding the maximum latency within the network.
- Ports on the router is the physical place where you can place the cable in the router.
- The number of ports on a router is the **radix** of the router

Different Network Topologies:

1. **N-Dimensional mesh / torus networks:**
    a. **Mesh Networks**: Configurations where each switch connects to a limited number of other switches. The network can be extended into multiple dimensions (e.g., 2D and 3D meshes) depending on the number of interconnected nodes per switch.
    b. Each switch has a small number of nodes connected to it
    c. Lets talk about 1D: If there are 2 nodes they can be connected by 1 link 1D mesh.
    d. If there are 4 nodes they can be connected by 4 links so that is 2D mesh
    e. If there are 6 nodes connected to 1 node then that is 3D mesh.
    f. **Dimensionality**: Each switch has direct links to 2n switches where n is the number of dimensions
    g. The cables are also bidirectional. They are not actually bidirectional but they are coaxial, which are cables that have the inner and outer cable where one of the cables carries data in 1 direction and the other cable carries data in the other direction.
    h. In the case of a mesh network we can create n dimensional meshes although 2 or 3 dimensions are the most common.
    i. Torus networks are just a modification of mesh networks. In torus networks we take the last nodes on each dimension and connect them back to the first node on the dimension to create a torus. These are also called wraparound links as they create a looped structure to the network. This helps us reduce the network diameter as seen below and also helps us improve redundancy.
    j. The total number of hops are all the cables that the message goes through from start node to destination, however, we don't always count the hop from the Nic to the switch. Switch and router is the term and it is used interchangeably.

k. Network diameter for Mesh Network (dimension X*Y where X and Y are the number of nodes in each dimension, more dimensions can be added)
    i. X+Y or (X-1)+(Y-1)
l. Network Diameter for Torus Network dimension dimension (X*Y, X and Y are number of nodes in each dimension but more can be added)
    i. (X+Y)/2

Electrical cables and Fiber Optic cable

- **Copper Cables:** Suitable for short distances within the same cabinet as signal strength diminishes over longer distances. Copper cables cannot carry the signal over very long distances without losing the signal. may not even be used over distances such as between cabinets.
- **Optical cables:** Though more costly, these are preferred for long-distance communication as they maintain signal integrity over greater distances.
- **Length**:
    - Assume the length of the cable has no effect on how long it takes for signal/data to be transmitted.

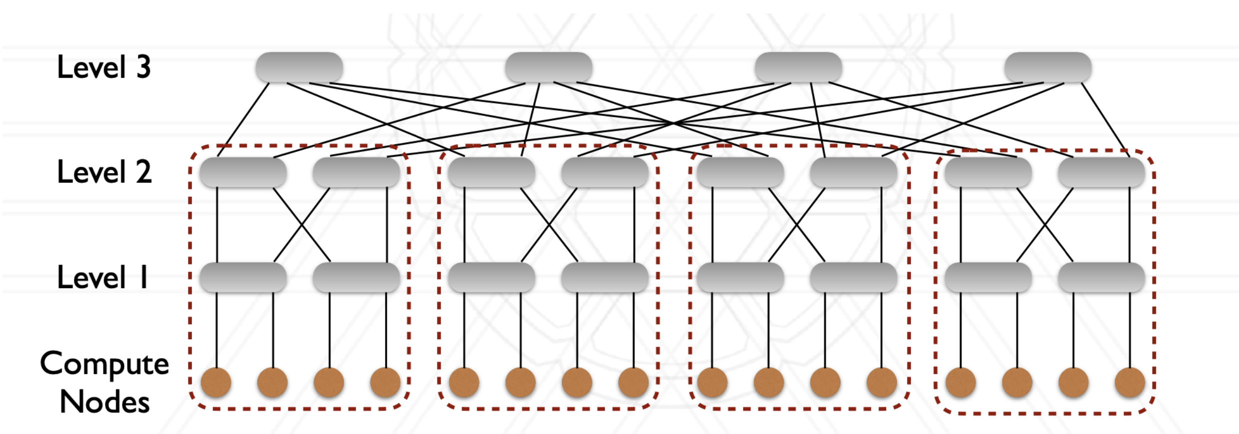
2. **Fat-tree Network:**

Motivation for this network comes from telecommunication networks

A Fat-tree network is a hierarchical network topology that is commonly used in HPC environments to interconnect compute nodes. It is called a "Fat-tree" because the "tree" branches out in terms of bandwidth as we move up the hierarchy, meaning that the higher-level switches have greater bandwidth capabilities to handle the cumulative traffic from below.

Key Components of a Fat-tree Network:

This is a diagram taken for the slides for this lecture, which I have described below to epxlain how a fat tree network works



1. **Compute Nodes**: These are connected to the lowest level of the network hierarchy (Level 1 in the diagram).
2. **Routers/Switches**: Devices that forward data packets between the compute nodes. In a Fat-tree, they are organized in levels, with Level 1 being the closest to the compute nodes and the highest level being the 'root' of the tree.
3. **Router Radix (k)**: The number of ports on a router/switch. In a Fat-tree, this number determines the scale of the network. For instance, if k=6, each router would have 6 ports.
4. **Pods**: A pod is a subsection of the Fat-tree network, typically containing $k/2$ switches at each level. The maximum number of pods is equal to the radix k. and each pod operates like a mini Fat-tree, with interconnectivity within the pod and uplinks to the higher-level switches.

How It Works:

- At the lowest level (Level 1), each switch connects to $k/2$ compute nodes, and the remaining $k/2$ ports are used to connect to the switches at Level 2.
- At Level 2, switches are connected downwards to the lower level switches within the same pod and upwards to the top-level switches (Level 3).
- At the top level (Level 3), the switches do not connect directly to compute nodes but serve as the backbone of the network, handling traffic between pods. They have connections coming in from multiple Level 2 switches across different pods.

Benefits:

- **Scalability**: Fat-tree topology can easily scale to accommodate more compute nodes by increasing the radix k and adding more levels or pods.
- **Fault Tolerance**: Due to multiple paths through the network, the Fat-tree is tolerant to individual node or switch failures.
- **Bandwidth**: As we move up the levels, the effective bandwidth increases because the upper-level switches aggregate traffic from multiple lower-level switches.